Journal of Agricultural Machinery

Homepage: https://jame.um.ac.ir

Research Article
Vol. 13, No. 2, 2023, p. 175-194

Maize Tassel Detection and Counting Using Deep Learning Techniques

Sh. Falahat Nejad Mahani[1], A. Karami[2*]

1- Ph.D. Student, Faculty of Physics, Shahid Bahonar University of Kerman, Kerman, Iran
2- Associate Professor, Faculty of Physics, Shahid Bahonar University of Kerman, Kerman, Iran
(*- Corresponding Author Email: azam.karami@gmail.com)
https://doi.org/10.22067/jam.2022.72477.1062

Received: 16 September 2021 Revised: 28 November 2021 Accepted: 02 January 2022 Available Online: 02 January 2022	**How to cite this article:** Falahat Nejad Mahani, Sh., & Karami, A. (2023). Maize Tassel Detection and Counting Using Deep Learning Techniques. *Journal of Agricultural Machinery*, *13*(2), 175-194. (in Persian with English abstract). https://doi.org/10.22067/jam.2022.72477.1062

Introduction

Maize is one of the most important cereal crops worldwide, providing staple food for people globally. Counting maize tassels provides essential information about yield prediction, growth status, and plant phenotyping, but traditional manual approaches are expensive and time-consuming. Recent developments in technology, including high-resolution RGB imagery acquired by unmanned aerial vehicles (UAVs) and advanced machine-learning techniques such as deep learning (DL), have been used to analyze genotypes, phenotypes, and crops.

In this study, we modified the YOLOv5s single-stage object detection technique based on a deep convolutional neural network and named it MYOLOv5s. We incorporated BottleneckCSP structures, Hardswish activation function, and two-dimensional spatial dropout layers to increase tassel detection accuracy and reduce overfitting. Our method's performance was compared with three state-of-the-art algorithms: Tasselnetv2+, RetinaNet, and Faster R-CNN. The results obtained from our proposed method demonstrate the effectiveness of MYOLOv5s in detecting and counting maize tassels.

Materials and Methods

The High-Intensity Phenotyping Site (HIPS) dataset was collected from the large field at the Agronomy Center for Research and Education (ACRE) of Purdue University, located in West Lafayette, Indiana, USA during the 2020 growing season. A Sony Alpha 7R-III RGB camera mounted on a UAV at a 20m altitude captured high-resolution orthophotos with a pixel resolution of 0.25 cm. The dataset consisted of two replications of 22 entries each for hybrids and inbreds, planted on May 12 using a two-row segment plot layout with a plant population of 30,000 per acre. The hybrids and inbreds in this dataset had varying flowering dates, ranging from 20 days between the first and last variety.

This article uses orthophotos taken on July 20th and 24th to train and test the proposed deep network "MYOLOv5s." These orthophotos were divided into 15 images (3670×2150) and then cropped to obtain 150 images (608 × 2048) for each date. Three modifications were applied to the original YOLOv5s to form MYOLOv5s: BottleneckCSP structures were added to the neck part of the YOLOv5s, replacing some C3 modules; two-dimensional spatial dropout layers were used in the defect layer; and the Hardswish activation function was utilized in the convolution structures. These modifications improved tassel detection accuracy. MYOLOv5s was implemented in the Pytorch framework, and the Adam algorithm was applied to optimize it. Hyper-parameters such as the number of epochs, batch size, and learning rates were also optimized to increase tassel detection accuracy.

Results and Discussion

In this study, we first compared the original and modified YOLOv5s techniques, and our results show that MYOLOv5s improved tassel detection accuracy by approximately 2.80%. We then compared MYOLOv5s performance to the counting-based approach TasselNetv2+ and two detection-based techniques: Faster R-CNN and RetinaNet. Our results demonstrated the superiority of MYOLOv5s in terms of both accuracy and inference time. The proposed method achieved an AP value of 95.30% and an RMSE of 1.9% at 84 FPS, making it about

1.4 times faster than the other techniques. Additionally, MYOLOv5s correctly detected the highest number of maize tassels and showed at least a 17.64% improvement in AP value compared to Faster R-CNN and RetinaNet, respectively. Furthermore, our technique had the lowest false positive and false negative values. The regression plots show that MYOLOv5s provided slightly higher fidelity counts than other methods.

Finally, we investigated the effect of score values on the performance of detection-based models and calculated the optimal values of hyperparameters.

Conclusion

1. The MYOLOv5s technique outperformed other state-of-the-art models in detecting maize tassels, achieving the highest precision, recall, and average precision (AP) values.
2. The MYOLOv5s method had the lowest root mean square error (RMSE) value in the error counting metric, demonstrating its accuracy in detecting and counting maize tassels.
3. We evaluated the correlation between predicted and ground-truth values of maize tassels using the R2 score, and for the MYOLOv5s method, the R2 score was approximately 99.28%, indicating a strong correlation between predicted and actual values.
4. The MYOLOv5s method performed exceptionally well in detecting tassels, even in highly overlapping areas. It accurately distinguished and detected tassels, regardless of their proximity or overlap with other objects.

When compared to the counting-based approach TasselNetv2+, our proposed MYOLOv5s method showed faster inference times. This suggests that the MYOLOv5s method is computationally efficient while maintaining accurate tassel detection capabilities.

Keywords: Deep learning, Image processing, Maize tassel, Object detection, UAV

نشریه ماشین‌های کشاورزی
https://jame.um.ac.ir

مقاله پژوهشی
جلد ۱۳، شماره ۲، تابستان ۱٤۰۲، ص ۱۹٤-۱۷۵

تشخیص و شمارش خودکار کاکل‌های ذرت با استفاده از آشکارساز مبتنی بر یادگیری عمیق

شهرزاد فلاحت نژاد ماهانی[۱]، اعظم کرمی[۲*]

تاریخ دریافت: ۱۴۰۰/۰۶/۲۵
تاریخ پذیرش: ۱۴۰۰/۱۰/۱۲

چکیده

در سال‌های اخیر استفاده از یادگیری عمیق در کشاورزی دقیق به‌منظور تشخیص و شمارش آفات و یا بیماری‌های گیاهان، سمپاشی هوشمند، تخمین سطح زیر کشت و نظارت بر روند رشد گیاهان جهت مقابله با عوامل بازدارنده و یا کاهش‌دهنده رشد و با هدف افزایش بهره‌وری محصولات کشاورزی به سرعت رو به افزایش است. در این مقاله، به طراحی الگوریتمی برگرفته از شبکه عصبی عمیق YOLOv5s جهت تشخیص و شمارش خودکار کاکل‌های گیاه ذرت پرداخته شده است. برای این منظور، از تصاویر اخذ شده توسط پهپاد از مزرعه ذرت در دو تاریخ متفاوت جهت آموزش و ارزیابی شبکه استفاده گردیده و با توجه به نوع و اندازه داده به اعمال تغییراتی در معماری و تابع فعال‌سازی الگوریتم اصلی YOLOv5s با هدف افزایش تعداد پارامترهای شبکه، کاهش بیش برازش و افزایش دقت تشخیص پرداخته شد و الگوریتم Modified YOLOv5s که به اختصار MYOLOv5s نام دارد به‌عنوان نسخه بهبودیافته YOLOv5s با قابلیت شناسایی و شمارش کاکل‌های ذرت با مقادیر ضریب تبیین R^2 ۹۹/۲۸ درصد و دقت متوسط (AP) ۹۵/۳۰ درصد حاصل شد. همچنین، عملکرد روش پیشنهادی به‌کار گرفته شده در این مقاله با الگوریتم‌های معتبر معرفی شده در این زمینه +TasselNetv2، Faster R-CNN و RetinaNet مقایسه گردید. نتایج به‌دست‌آمده نشان می‌دهد که مقادیر ضریب تبیین برای این سه شبکه به‌ترتیب ۷۷/۸۶، ۸۶/۸۳ و ۹۵/۵۳ درصد می‌باشد. همچنین برای الگوریتم‌های Faster R-CNN و RetinaNet مقادیر دقت متوسط ۷۶/۹۹ و ۷۷/۶۶ درصد به‌دست آمد. این نتایج نشان می‌دهند که الگوریتم MYOLOv5s حداکثر مقادیر ضریب تبیین R^2 و دقت متوسط (AP)، دقت (Precision) و یادآوری (Recall) را دارد که بیانگر کارایی بالای روش پیشنهادی در تشخیص کاکل ذرت است. شایان ذکر است MYOLOv5s با دارا بودن سرعت پردازش ۸۴ فریم بر ثانیه سریع‌ترین روش در تشخیص کاکل ذرت محسوب می‌گردد.

واژه‌های کلیدی: پردازش تصویر، پهپاد، تشخیص شی، کاکل ذرت، یادگیری عمیق

مقدمه

ذرت به‌عنوان یکی از مهم‌ترین غلات در فراهم آوردن تغذیه بخش کثیری از جمعیت کره زمین به‌شمار می‌رود. همچنین، به‌طور گسترده‌ای به‌عنوان خوراک دام، سوخت زیستی و مواد اولیه در صنعت نیز کاربرد دارد (Tagne, Feujio, & Sonna, 2008). از این رو، برداشت حداکثری این محصول حائز اهمیت می‌باشد. این امر نیازمند کاهش و یا ترجیحاً حذف عوامل آسیب‌زا به گیاهان نظیر خاک نامرغوب، آفات، بیماری‌ها، خورده شدن بذر توسط حشرات و پرندگان و غیره می‌باشد. از این رو نظارت مستمر بر روند رشد و تعداد گیاهان ذرت یک مزرعه به‌منظور مقابله با این‌گونه اثرات آسیب‌زا در صورت مشاهده تغییر در رشد امری ضروری است (Parihar *et al.*, 2011; Lu, Cao, Xiao, Zhuang, & Shen, 2017). علاوه براین جمعیت یک گیاه به‌عنوان یک شاخص اساسی در بررسی خواص فنوتایپینگ آن نیز به‌شمار می‌روند. نظارت بر گیاه ذرت در زمان گل‌دهی آن نیز نقش مهمی در تعیین بهره‌وری این محصول دارد زیرا اگر گیاهان خیلی زود گل بدهند محصول قبل از پایان فصل رشد بالغ می‌شود و فرصت جذب و استفاده از مقادیر زیادی انرژی نور برای فوتوسنتز را از دست می‌دهند و چنانچه خیلی دیر گل بدهند ممکن است محصول قبل از آماده شدن برای برداشت با تغییر فصل از بین برود. لذا، حتی مقادیر اندک تنش می‌تواند عملکرد را به میزان قابل‌توجهی تغییر دهد (Alzadjali *et al.*, 2021). در گذشته این نظارت از طریق اشخاص انجام می‌گرفت که امری طاقت‌فرسا، زمان‌بر و غیردقیق بود. اما اکنون برای فایق آمدن بر این مشکل، روش‌هایی نوین برای شمارش

۱ و ۲- به‌ترتیب دانشجوی دکترا و دانشیار دانشکده فیزیک، دانشگاه شهید باهنر کرمان، کرمان، ایران
(*- نویسنده مسئول: Email: azam.karami@gmail.com)
https://doi.org/10.22067/jam.2022.72477.1062

خودکار محصولات زراعی مبتنی بر هوش مصنوعی و پردازش تصویر در سالیان اخیر پیشنهاد شده است.

از جمله به‌کارگیری روش‌های هوش مصنوعی و پردازش تصویر در حوزه کشاورزی می‌توان به شناسایی علف‌های هرز (Parihar *et al.*, 2011)، تشخیص بیماری‌ها (Gómez-Flores, Garza-Saldaña, & Varela-Fuentes, 2019; Pourreza, Lee, Etxeberria, & Banerjee, 2015)، شمارش کاکل‌های ذرت (Lu *et al.*, 2017)، شناسایی کاکل‌های ذرت (Liu *et al.*, 2020)، شمارش برگ (Ubbens, Cieslak, Prusinkiewicz, & Stavness, 2018)، تشخیص نهال‌ها (Quan *et al.*, 2019) و شناسایی و شمارش سورگوم (Ghosal *et al.*, 2019) اشاره کرد. تمامی مقالات نامبرده با طراحی الگوریتم‌های مختلف مبتنی بر هوش مصنوعی و پردازش تصویر توانسته‌اند عمل شمارش و یا تشخیص شی مورد نظر را به‌صورت خودکار انجام داده و با کاهش زمان و افزایش قدرت شمارش و یا تشخیص گام بزرگی در جهت کشاورزی هوشمند بردارند. هرچند کاهش زمان تشخیص و افزایش بیشتر دقت هنوز یک مسئله باز محسوب می‌گردد و هر پیشرفتی در این امر شایسته تقدیر است.

الگوریتم‌های هوش مصنوعی به‌کار رفته در مقالات ذکر شده در دو گروه روش‌های کلاسیک و یادگیری عمیق[1] جهت شمارش و یا تشخیص شی مورد نظر قابل بحث هستند. از جمله روش‌های کلاسیک می‌توان به تشکیل تصویر دودویی، روش‌های مورفولوژی، لبه‌یابی و روش‌های یادگیری ماشین اشاره کرد. در این روش‌ها استخراج ویژگی‌های با ارزش تصویر که به‌منظور کاهش بار محاسباتی الگوریتم باید انجام گیرد به‌صورت دستی و مجزا از الگوریتم طراحی‌شده جهت تشخیص و یا شمارش شی باید انجام گیرد اما در یادگیری عمیق این ویژگی‌ها در ساختاری سراسری[2] توسط لایه‌های شبکه عمیق به‌صورت خودکار استخراج می‌شوند (Ongsulee, 2017). از این رو، استفاده از الگوریتم‌های یادگیری عمیق در تشخیص و یا شمارش اشیا به نتایج بهتری منتهی می‌گردد. الگوریتم‌های یادگیری عمیق به‌کار رفته برای این منظور خود نیز به دو دسته الگوریتم‌های مبتنی بر شمارش و تشخیص قابل دسته‌بندی هستند. الگوریتم‌های یادگیری عمیق مبتنی بر شمارش با به‌کارگیری یک الگوریتم رگرسیون مناسب داده‌های مسئله و ایجاد رابطه‌ای معنادار بین متغیرهای وابسته و تعداد اشیا موجود در تصویر به‌طور مستقیم به پیش‌بینی تعداد شی مورد نظر در سایر تصاویر می‌پردازند حال آن‌که در روش‌های مبتنی بر تشخیص از یک شبکه عمیق برای تشخیص و شمارش اشیا در تصویر استفاده می‌شود. این عمل از طریق رسم کادرهای محصورکننده با امتیاز معین در اطراف شی و یا رسم نقطه در مرکز آن انجام می‌گیرد (Lempitsky & Zisserman, 2010).

اخیراً، تعدادی مدل یادگیری عمیق نیز برای تشخیص و شمارش کاکل‌های ذرت پیشنهاد شده‌اند. الگوریتم‌های بر پایه شمارش TasselNet (Lu *et al.*, 2017)، TasselNetv2 (Xiong *et al.*, 2019) و +TasselNetv2 (Lu & Cao, 2020) از جمله مدل‌های مورد استفاده برای شمارش هستند. همان‌گونه که گفته شد این مدل‌ها از یک الگوریتم رگرسیون جهت پیش‌بینی تعداد کاکل‌های ذرت استفاده می‌کنند و +TasselNetv2 به‌عنوان نسخه بهبودیافته دو الگوریتم TasselNet و TasselNetv2 قادر خواهد بود کاکل‌های ذرت موجود در تصاویر ابعاد بالا را در زمان کم بشمارد. شبکه‌های Faster R-CNN (Liu *et al.*, 2020) و RetinaNet (Zou, Lu, Li, Liu, & Cao, 2020) برای تشخیص و شمارش کاکل‌ها به‌کار گرفته شده‌اند. در مدل مبتنی بر Faster R-CNN جهت این امر از دو شبکه پشتیبان VGGNet[3] و ResNet[4] استفاده شده است و با بهبود اندازه لنگرها[5] مدل شناسایی کاکل ذرت طراحی شده است. در طراحی مدل بر پایه RetinaNet نیز از ResNet34 به‌عنوان شبکه پشتبان مدل تشخیص کاکل ذرت استفاده شده است. از مقایسه این روش‌ها با هم در (Zou *et al.*, 2020) این نتیجه حاصل شد که مدل‌های استفاده شده جهت شمارش نسبت به روش‌های بر پایه تشخیص سریع‌تر عمل می‌کنند و توانایی بیشتری در شمارش کاکل‌های با همپوشانی بالا دارند اما علی‌رغم روش‌های مبتنی بر تشخیص قادر به تعیین دقیق موقعیت مکانی و اندازه هر کاکل نمی‌باشند.

بدین جهت، برای هرچه کارآمدتر بودن روش‌های یادگیری عمیق در شناسایی کاکل‌های ذرت طراحی الگوریتمی مبتنی بر تشخیص که دارای سرعت پردازش بالا نیز باشد برای امر شمارش گیاهان بسیار حائز اهمیت می‌باشد. از این رو، در این مقاله به طراحی شبکه MYOLOv5s به‌عنوان نسخه بهبودیافته الگوریتم مبتنی بر تشخیص YOLOv5s (Jocher *et al.*, 2020) برای تحلیل داده‌های کاکل ذرت پرداخته شده است. این الگوریتم به‌صورت تخصصی با توجه به نوع مجموعه داده و ویژگی‌های کاکل‌های ذرت در آن (شکل، اندازه، رنگ، بافت و غیره) جهت شناسایی این گیاه به‌صورت اختصاصی طراحی شده است. در ادامه به نوآوری‌های انجام گرفته در طراحی الگوریتم MYOLOv5s اشاره خواهد شد:

۱- جایگزین کردن برخی ساختارهای C3 در قسمت گردن[6] شبکه

1- Deep Learning
2- End-to-End

3- Visual Geometry Group Neural Network
4- Residual Neural Network
5- Anchor
6- Neck

YOLOv5s با ساختار BottleneckCSP به‌منظور افزایش قابلیت یادگیری شبکه

۲- استفاده کردن از تابع فعال‌سازی Hardswish جهت افزایش قابلیت یادگیری شبکه

۳- قرار دادن لایه‌های حذف تصادفی مکانی[1] در قسمت سر[2] شبکه به‌منظور جلوگیری از بیش‌برازش

۴- بهینه کردن ابر پارامترهای[3] الگوریتم به‌منظور افزایش دقت تشخیص کاکل‌های ذرت

اعمال اصلاحات فوق ضمن دستیابی به دقت بالای تشخیص کاکل‌های ذرت در مدت زمان کم شرایط را برای عملکرد مطلوب شبکه با تعداد داده آموزش کم فراهم آورده که خود با توجه به هزینه‌های بالای تشکیل پایگاه داده دستاورد مهمی محسوب می‌گردد. همچنین، جهت بررسی بیشتر عملکرد این مدل نتایج حاصل از آن با سه شبکه TasselNetv2+، Faster R-CNN، RetinaNet مقایسه شده و برتری شبکه طراحی‌شده در تشخیص و شمارش کاکل‌های ذرت در کوتاه‌ترین زمان به اثبات رسیده است.

مواد و روش‌ها

مجموعه داده HIPS[4]

در این مقاله، برای طراحی مدل MYOLOv5s از داده‌های تصویری تهیه شده توسط بخش تحقیقات کشاورزی دانشگاه پردو[5]، ایندیانا، آمریکا استفاده شده است. این داده‌ها متشکل از تصاویر ۲۲ ردیف کاکل ذرت اینبرد[6] و ۲۲ ردیف کاکل ذرت هیبرید[7] هستند که با جمعیت ۳۰۰۰۰ گیاه بر هکتار در تاریخ ۱۲ ماه می سال ۲۰۲۰ میلادی کاشته شده‌اند. تصویربرداری در طول فصل رشد ذرت در تاریخ‌های ۲۰ و ۲۴ جولای از ۱۵ ناحیه با استفاده از دوربین رنگی Sony Alpha 7R-lll نصب‌شده بر روی پهپاد DJI Matrice M600 Pro از ساعت ۱۰:۰۰ الی ۱۲:۰۰ صبح که نور مناسب برای تصویربرداری وجود دارد و از ارتفاع ۲۰ متری انجام گرفت. این دوربین دارای رزولوشن ۴۲/۴ مگاپیکسل است. سپس تصاویر خام رنگی پردازش شدند و به ۱۵ اورتوفوتو[8] (Habib, Kim, & Kim, 2007) رنگی بسیار با کیفیت برای هر تاریخ با اندازه میانگین ۳۶۷۰×۲۱۵۰ و رزولوشن مکانی ۰/۲۵ سانتی‌متر تبدیل شدند.

اطلاعات فنوتیپی از جمله شمارش و تاریخ گل‌دهی با بازرسی‌های بصری در مزرعه در طول مدت رشد نیز ثبت شده است. در شکل ۱ نمونه‌ای از اورتوفوتوهای تاریخ‌های ۲۰ و ۲۴ جولای نشان داده است.

همان‌گونه که از تصاویر پیداست در تاریخ ۲۴ جولای در مقایسه با تاریخ ۲۰ جولای ذرت‌ها رشد بیشتری داشته‌اند، هم‌پوشانی بیشتری دارند، خاک خشک است و تصویربرداری در شرایط جوی متفاوت از تاریخ ۲۰ جولای انجام گرفته است.

به‌منظور آماده‌سازی مجموعه داده جهت استفاده برای طراحی مدل، هر تصویر اورتوفوتو باید به تصاویر کوچک‌تر برش داده شود. برش تصاویر با اهداف استفاده بهینه از این تصاویر، افزایش دقت الگوریتم و کاهش مدت زمان پردازش داده انجام می‌گیرد و با توجه به نوع داده باید به گونه‌ای باشد که تغییر چندانی در اشیا موجود در آن ایجاد نکند، توسط الگوریتم در نظر گرفته شده قابل‌پردازش باشد، عملکرد الگوریتم را بهبود بخشد و متناسب با داده‌های آزمون تعیین شده توسط مرکز تحقیقات کشاورزی دانشگاه پردو آمریکا در مجموعه داده باشد. داده آزمون تعیین شده توسط این مرکز برای هر تصویر اورتوفوتو شامل دو ردیف سمت چپ هر تصویر می‌باشد و ۱۰ درصد کل تصاویر را تشکیل می‌دهد. این تصاویر متشکل از کاکل‌های ذرت اینبرد و هیبرید بوده و با اندکی تفاوت برای ۳۰ تصویر مختلف اندازه‌ای در حدود ۶۰۰×۲۱۵۰ پیکسل دارد. با توجه به این‌که جهت حصول نتیجه بهتر در زمان آموزش اندازه تصاویر آموزش بهتر است که نزدیک به تصاویر آزمون انتخاب شود لذا اندازه ۶۰۸×۲۰۴۸ به‌عنوان اندازه مناسب برش تصاویر در این الگوریتم انتخاب شد و بدین طریق هر تصویر به ۱۰ تصویر کوچک‌تر برش داده شد. برش هر تصویر اورتوفوتو در این کار به گونه‌ای انجام گرفته است که کاکل‌های گیاه ذرت برش داده نشوند و الگوریتم با دقت بیشتری و بدون تشخیص نادرست قادر به شناسایی درست و دقیق کاکل‌های ذرت باشد. سپس تصاویر از طریق رسم کادرهای محصورکننده[9] اطراف کاکل‌های ذرت برچسب زده شدند و بدین طریق ۴۰۴۶ کادر محصورکننده برای این مجموعه داده به‌دست آمد. پس از آن با توجه به استاندارد موجود ۱۰ درصد باقی‌مانده تصاویر (کل تصویر به‌جز دو ردیف آخر از سمت چپ) برای اعتبارسنجی و مابقی جهت آموزش در نظر گرفته شدند.

شرح مدل

YOLO[10] (Redmon, Divvala, Girshick, & Farhadi, 2016) از جمله شبکه‌های عصبی پیچشی[11] (Albawi,

1- Spatial-Dropout
2- Head
3- Hyper-parameter
4- High Intensity Phenotyping Sites
5- Purdue University
6- Inbred
7- Hybrid
8- Orthophoto
9- Bounding Box
10- You Look Only Once
11- Convolutional Neural Network

(Mohammed, & Al-Zawi, 2017) و الگوریتمی تک مرحله‌ای و یک‌پارچه برای تشخیص اشیا است که از طریق رسم کادرهای محصورکننده با امتیاز معین، که بیان‌کننده احتمال وجود شی در کادر است، به شناسایی آن شی و برچسب مربوط به آن می‌پردازد. این الگوریتم از آن جهت تک مرحله‌ای نام گرفته که عملیات شناسایی شی و اختصاص برچسب به آن را به یکباره انجام می‌دهد. اولین نسخه آن در سال ۲۰۱۶ منتشر شد و پس از آن YOLOv2 (Redmon & Farhadi, 2017)، YOLOv3 (Redmon & Farhadi, 2018)، YOLOv4 (Bochkovskiy, Wang, & Liao, 2020)، YOLOv5 و PP-YOLO (Long *et al.*, 2020) با هدف بهبود دقت و سرعت تشخیص نسخه‌های پیشین ارائه شدند.



الف) تصویر اخذ شده در ۲۰ جولای ۲۰۲۰

a) Captured orthophotos on 2020/07/20



ب) تصویر اخذ شده در ۲۴ جولای ۲۰۲۰

b) Captured orthophotos on 2020/07/24

شکل ۱- تصاویر اورتوفوتو اخذ شده از مزرعه ذرت در تاریخ‌های ۲۰ و ۲۴ جولای

Fig.1. Captured orthophotos from corn farm on July 20 & 24

YOLOv5 خود نیز شامل چهار الگوریتم YOLOv5s، YOLOv5m، YOLOv5l و YOLOv5x با عمق‌های مختلف است. هرچه شبکه‌ای عمیق‌تر باشد تعداد لایه‌های بیشتری جهت آموزش شبکه به‌کار رفته و در نتیجه شبکه ویژگی‌های بیشتری را می‌آموزد و کندتر عمل می‌کند. در این بین، جهت طراحی الگوریتم تشخیص و شمارش کاکل‌های ذرت MYOLOv5s با توجه به ظرفیت سیستم پردازش داده در ابتدا از الگوریتم YOLOv5s به‌عنوان کم‌عمق‌ترین و سریع‌ترین الگوریتم از بین الگوریتم‌های نامبرده استفاده می‌شود و سپس با اعمال پاره‌ای تغییرات بر روی آن، الگوریتم بهبودیافته مختص شناسایی انواع کاکل‌های ذرت ایجاد می‌گردد.

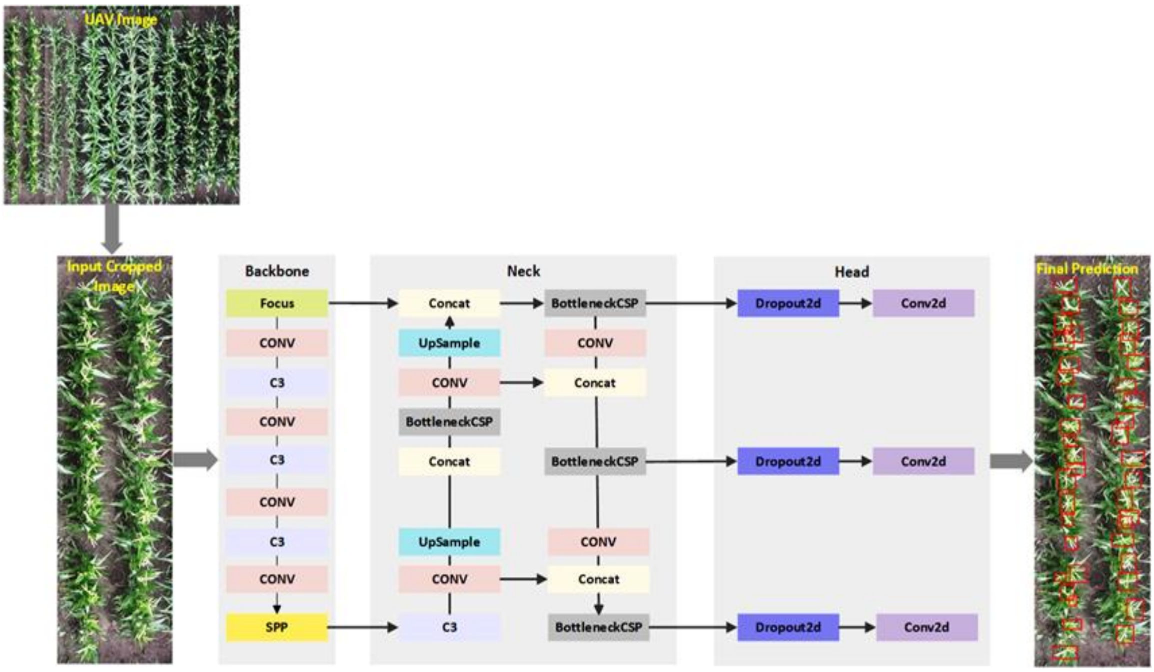
معماری الگوریتم

الگوریتم MYOLOv5s از سه قسمت شبکه پشتبان، گردن و سر تشکیل شده است. معماری این الگوریتم در شکل ۲ نشان داده شده است. در ادامه جزئیات این معماری شرح داده خواهد شد.

ساختارهای تمرکز[1]، پیچشی[2] (CONV)، C3 و ادغام هرم مکانی[3] (SPP) (He, Zhang, Ren, & Sun, 2015) شبکه پشتبان شبکه را بر مبنای شبکه مرتبه جزئی متقاطع[4] (CSPNet) (Wang *et al.*, 2020) شکل می‌دهند. در این قسمت، ابتدا به‌منظور تسریع الگوریتم، یک نقشه ویژگی[5] از تصویر ورودی بر مبنای جابه‌جایی پیکسل‌ها[6] (Shi *et al.*, 2016) در اندازه نصف ابعاد مکانی تصویر ورودی و ۱۲ باند توسط ساختار تمرکز ایجاد شده و به ساختار پیچشی وارد می‌شود. سپس، تصویر ایجاد شده به ساختار C3 (Jocher *et al.*, 2020) که مشابه با BottleneckCSP (Wang *et al.*, 2020) اما با سه لایه پیچشی است وارد می‌شود تا از طریق تقسیم باندها توسط این ساختار شرایط برای یادگیری تعداد بیشتری ویژگی توسط الگوریتم فراهم آید. در نهایت، در این قسمت، از ساختار SPP برای ادغام ویژگی‌ها با مقیاس‌های مختلف و افزایش قابلیت الگوریتم بهره برده می‌شود.

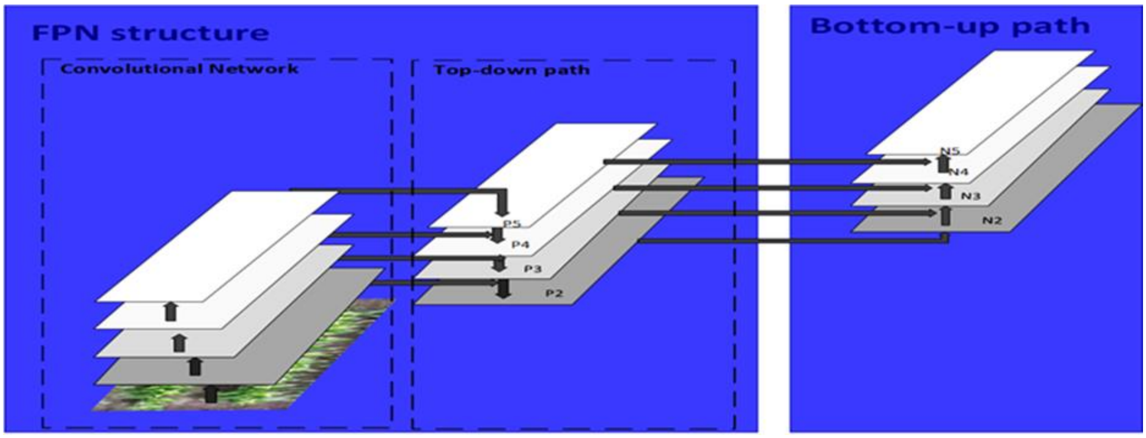
1- Focus
2- Convolutional

3- Spatial Pyramid Pooling
4- Cross Stage Partial Network
5- Feature Map
6- Pixel Shuffling



شکل ۲– معماری الگوریتم MYOLOv5s

Fig.2. MYOLOv5s architecture



شکل ۳– ساختار PANet استفاده شده در گردن مدل

Fig.3. PANet structure used in the neck part of the model

گردن در الگوریتم MYOLOv5s از نوع شبکه تجمیع مسیر[1] (PANet) (Liu, Qi, Qin, Shi, & Jia, 2018) است. همان‌گونه که در شکل ۳ پیداست شبکه PANet با اضافه کردن یک مسیر پایین به بالا به شبکه ویژگی هرمی[2] (FPN) (Lin *et al.*, 2017)، که خود متشکل از یک مسیر بالا به پایین و یک ساختار پیچشی می‌باشد، ایجاد شده است. ساختارهای اتصال[3] (Concat)، نمونه‌برداری افزایشی[4]، پیچشی، C3 و BottleneckCSP اجزای تشکیل‌دهنده PANet در نسخه بهبودیافته MYOLOv5s هستند که با هدف افزایش قابلیت الگوریتم در شناسایی اشیا با اندازه و مقیاس‌های مختلف کنار یکدیگر قرار گرفته‌اند. قراردهی

1- Path Aggregation Network
2- Feature Pyramid Network
3- Concatenation
4- Up-Sampling

BottleneckCSP به‌عنوان ساختاری با عملکردی مشابه با C3 اما با تعداد پارامتر بیشتر در کنار سایر اجزای گردن در این الگوریتم به‌عنوان بخش نوآورانه طرح سبب افزایش تعداد پارامترهای الگوریتم به‌منظور یادگیری ویژگی‌های بیشتر می‌گردد و با توجه به اندازه کوچک مجموعه داده HIPS و شکل خاص کاکل‌های ذرت نتایج بهتر تشخیص شی را نتیجه می‌دهد.

قسمت سر الگوریتم آخرین قسمت در معماری MYOLOv5s است که توسط لایه تشخیص متشکل از لایه‌های پیچشی به تشخیص نهایی اشیا در تصویر ورودی می‌پردازد. با توجه به این‌که الگوریتم YOLOv5s یک الگوریتم پیچشی است و از طریق ویژگی‌ها آموزش می‌یابد و پیکسل‌های مجاور در نقشه ویژگی تصویر بسیار همبسته‌اند. لذا، افزودن لایه‌های حذف تصادفی مکانی (Tompson, Goroshin, Jain, LeCun, & Bregler, 2015) قبل از لایه‌های پیچشی قسمت سر در الگوریتم بهبودیافته MYOLOv5s به‌عنوان بخشی از نوآوری طرح سبب صفر شدن تصادفی نقشه ویژگی‌های دوبعدی شده، میزان همبستگی بین پیکسل‌ها را کاهش می‌دهد، از بیش‌برازش[1] (Hawkins, 2004) جلوگیری می‌کند و دقت تشخیص الگوریتم را بهبود می‌بخشد. لازم به ذکر است، با توجه به کوچک بودن مجموعه داده، متفاوت بودن شکل کاکل‌های ذرت، محدود بودن برخی ویژگی‌ها و کم‌عمق بودن الگوریتم، استفاده از تابع فعال‌سازی Hardswish (Howard *et al.*, 2019) به جای SiLU (Elfwing, Uchibe, & Doya, 2018) در الگوریتم بهبودیافته MYOLOv5s نوآوری دیگر طرح محسوب شده و منجر به یادگیری ویژگی‌های بیشتر توسط الگوریتم و عملکرد بهتر در تشخیص کاکل‌های ذرت با اشکال متفاوت می‌گردد.

جزئیات آموزش و تشخیص شی در الگوریتم

آموزش در الگوریتم MYOLOv5s مانند سایر شبکه‌های عصبی پیچشی از طریق روش پس‌انتشار خطا[2] (Leung & Haykin, 1991) انجام می‌گیرد. در این روش، ابتدا داده‌های آموزش به شبکه داده می‌شوند و سپس ویژگی‌های آن‌ها در حرکتی رو به جلو[3] لایه به لایه شبکه عمیق را طی می‌کنند تا جهت شناسایی شی آموزش داده شوند. آموزش به گونه‌ای انجام می‌پذیرد که یک تابع خطا حداقل گردد. پس از آن جهت ارزیابی عملکرد شبکه آموزش‌یافته، شبکه با داده‌های اعتبارسنجی بررسی می‌شود و در صورتی‌که خطای آموزش زیاد باشد مجدداً داده‌ها از طریق لایه‌ها برگشت داده می‌شوند و تا زمانی‌که نتیجه‌ای نزدیک‌تر به خروجی واقعی و با خطای کم‌تر دست آید و یا تعداد معینی تکرار سپری شود مراحل قبل تکرار می‌شوند. در پایان برای سنجش بی‌طرفانه الگوریتم آموزش‌دیده نهایی از داده‌های آزمون استفاده می‌شود. در ادامه طریقه شناسایی و شمارش شی و تابع خطای استفاده شده در این شبکه شرح داده خواهند شد.

شناسایی و شمارش شی

شناسایی شی توسط این الگوریتم در سه مرحله انجام می‌گیرد. در ابتدا، ویژگی‌های با ارزش تصویر ورودی توسط شبکه پشتیبان شبکه استخراج می‌شوند. سپس، در قسمت گردن شبکه این ویژگی‌ها با هم آمیخته شده و سه نقشه ویژگی به‌صورت S×S شبکه مشبک با گام‌های ۸، ۱۶ و ۳۲ جهت افزایش قابلیت شبکه در شناسایی اشیا کوچک، متوسط و بزرگ شکل می‌دهد. در آخر، تشخیص شی در قسمت سر شبکه از طریق در نظر گرفتن سه کادر محصورکننده برای هر سلول شبکه مشبک و نسبت دادن بهترین کادر به شی انجام می‌گیرد. تعداد کادرهای تعیین‌شده در انتها تعداد اشیا را مشخص می‌کنند و اندازه هر یک نمایان‌کننده اندازه شی مورد نظر است.

تابع خطا

تابع خطا الگوریتم‌های YOLO در طول آموزش ترکیبی از امتیاز شی بودن[4] (L_{conf})، امتیاز احتمال کلاس[5] (L_{cl}) و امتیاز رگرسیون کادر محصورکننده[6] (L_{GIoU}) است (Redmon *et al.*, 2016). امتیاز شی بودن، احتمال وجود شی درون کادر محصورکننده، امتیاز احتمال کلاس احتمال برچسب درست زدن برای هر شی و امتیاز رگرسیون کادر محصورکننده میزان هم‌پوشانی کادر محصورکننده پیش‌بینی‌شده و کادر حقیقی را تعیین می‌کند. در MYOLOv5s خطای امتیاز شی و احتمال کلاس از طریق آنتروپی متقاطع دودویی[7] همراه با تابع خطای منطقی[8] و خطای کادر محصورکننده توسط اشتراک پیرامون اجتماع کلی[9] (GIoU) (Rezatofighi *et al.*, 2019) به‌دست می‌آید. توابع خطای نامبرده در روابط (۱) تا (۴) نشان داده شده‌اند:

$$L_{conf} = -\sum_{i=0}^{S^2}\sum_{j=0}^{B} I_{ij}^{obj}\left[\hat{C}_i \log(C_i) + (1-\hat{C}_i)\log(1-C_i)\right] - \lambda_{noobj}\sum_{i=0}^{S^2}\sum_{j=0}^{B} I_{ij}^{noobj}\left[\hat{C}_i \log(C_i) + (1-\hat{C}_i)\log(1-C_i)\right] \quad (۱)$$

$$L_{cl} = -\sum_{i=0}^{S^2} I_{ij}^{obj} \sum_{c\in class}\left[\hat{P}_i(c)\log(P_i(c)) + (1-\hat{P}_i(c))\log(1-P_i(c))\right] \quad (۲)$$

1- Overfitting
2- Back Propagation of Error
3- Feed Forward

4- Objectness Score
5- Class Probability Score
6- Bounding Box Regression Score
7- Binary Cross-Entropy
8- Logits Loss Function
9- Generalized Intersection Over Union

$$L_{GIoU}=1-\frac{\left|b\cap b^{gt}\right|}{\left|b\cup b^{gt}\right|}-\frac{\left|G\setminus\left(b\cup b^{gt}\right)\right|}{\left|G\right|} \quad (۳)$$

$$LOSS=L_{GIoU}+L_{conf}+L_{cl} \quad (۴)$$

در این روابط S^2 بیانگر شبکه مشبک S×S و B کادرهای محصورکننده داخل هر سلول است. λ_{noobj} فاکتور وزن می‌باشد و I_{ij}^{obj} و I_{ij}^{noobj} در صورت وجود شی داخل به‌ترتیب کادر مقادیر ۱ و ۰ را اتخاذ می‌نمایند و در صورت نبود شی برعکس این امر اتفاق می‌افتد. $\hat{C}_i$ و C_i نیز نمایانگر امتیاز قطعیت کادرهای محصورکننده پیش‌بینی شده و حقیقی می‌باشند. احتمال کادر پیش‌بینی شده و حقیقی نیز برای هر کلاس با $\hat{P}_i(c)$ و $P_i(c)$ نمایش داده می‌شوند. b و b^{gt} کادرهای محصورکننده پیش‌بینی و حقیقی را نشان می‌دهند و G کوچک‌ترین کادری است که b و b^{gt} را احاطه می‌کند.

جزئیات پردازش

در این کار، تمامی پردازش‌ها در سرویس گوگل کولب[۱] (Bisong, 2019) با ۲۵ گیگابایت حافظه رم[۲] (RAM) و واحد پردازش گرافیکی[۳] (GPU) مدل Tesla P100 با ۱۶ گیگابایت حافظه انجام گرفت. MYOLOv5s مبتنی بر کتابخانه پایتورچ[۴] (Paszke *et al.*, 2019) پیاده‌سازی شده است. این مدل از بهینه‌ساز آدام[۵] جهت به‌روزرسانی وزن‌ها در طول آموزش استفاده می‌کند. تعداد تکرارها[۶] برای این شبکه ۲۰۰ و اندازه دسته[۷] جهت تعیین تعداد تصاویری که در قالب ورودی در هر تکرار به شبکه ارائه می‌گردد ۱۶ در نظر گرفته شده‌اند.

ارزیابی مدل

پس از آن‌که شناسایی کاکل‌ها توسط شبکه انجام گرفت، مجموعه‌ای از کادرهای محصورکننده که هر یک امتیازی بین صفر و یک را دارا می‌باشند به‌دست می‌آیند. با در نظر گرفتن یک مقدار آستانه برای امتیاز و اشتراک پیرامون اجتماع (IoU) که بیان‌کننده مقدار همپوشانی کادر محصورکننده حقیقی و تشخیص‌داده شده است، تعداد کادرهایی که به‌درستی توسط شبکه بر روی کاکل‌های ذرت زده شده‌اند[۸] (TP)، تعداد کادرهایی که به اشتباه زده نشده‌اند[۹] (FN) و تعداد کادرهایی که اشتباهاً در محل‌های دیگر زده شده‌اند[۱۰] (FP) قابل محاسبه هستند. با استفاده از این مقادیر معیارهای ارزیابی (۵) تا (۸) محاسبه می‌شوند (Sokolova, Japkowicz, & Szpakowicz, 2006):

$$Precision=\frac{TP}{TP+FP} \quad (۵)$$

$$Recall=\frac{TP}{TP+FN} \quad (۶)$$

$$F1=\frac{TP}{TP+\frac{1}{2}\left(FP+FN\right)} \quad (۷)$$

$$Accuracy=\frac{TP+TN}{TP+FP+TN+FN} \quad (۸)$$

در روابط ذکر شده دقت[۱۱] بیانگر تعداد تشخیص‌های انجام‌گرفته صحیح، یادآوری[۱۲] بیانگر تعداد کاکل‌هایی است که به اشتباه تشخیص داده نشده‌اند، معیار امتیاز F1 میانگین متوازن دقت و یادآوری و صحت[۱۳] نسبت بین تشخیص‌های صحیح به کل تشخیص‌ها است. معیار امتیاز F1 و صحت عملکرد شبکه را در حالت کلی نشان می‌دهند.

با در نظر گرفتن سطح زیر منحنی دقت-یادآوری دقت میانگین[۱۴] به‌دست می‌آید (Zhu, 2004):

$$AP=\int_0^1 p(r)dr \quad (۹)$$

در این رابطه $p(r)$ مقدار بیشینه دقت در مقدار یادآوری r است.

به‌منظور مقایسه نتایج به‌دست آمده توسط روش‌های مبتنی بر تشخیص با روش‌های مبتنی بر شمارش، تعدادی معیار که در روابط (۱۰) و (۱۱) نشان داده شده‌اند جهت ارزیابی عملکرد روش‌های مختلف در شمارش کاکل‌های ذرت در این مقاله استفاده شده‌اند (Chicco, Warrens, & Jurman, 2021):

$$RMSE=\sqrt{\frac{1}{N}\sum_1^N\left(g_j-c_j\right)^2} \quad (۱۰)$$

$$R^2=1-\frac{\sum_j\left(g_j-c_j\right)^2}{\sum_j\left(g_j-\bar{g}\right)^2} \quad (۱۱)$$

ضریب تبیین R^2 نشان‌دهنده میزان همبستگی بین تعداد حقیقی کاکل‌های ذرت و تعداد به‌دست‌آمده آن‌ها توسط الگوریتم و معیار

1- Google Colab
2- Random-Access Memory
3- Graphics Processing Unit
4- Pytorch
5- Adam
6- Epoch
7- Batch Size
8- True Positive
9- False Negative
10- False Positive
11- Precision
12- Recall
13- Accuracy
14- Average Precision (AP)

خطای جذر میانگین مربعات[1] (RMSE) بیان‌کننده اختلاف بین این دو دسته مقدار است.

در این روابط، N تعداد تصاویر آزمون، c_j و g_j به‌ترتیب تعداد کاکل‌های ذرت به‌دست‌آمده توسط الگوریتم و تعداد حقیقی آن‌ها برای تصویر j ام و $\bar{g}$ میانگین تعداد حقیقی کاکل‌های ذرت برای تصاویر آزمون است.

تعداد فریم بر ثانیه[2] (fps) معیار دیگری است که برای ارزیابی سرعت عملکرد الگوریتم در این مقاله استفاده شده است.

نتایج و بحث

در این قسمت ابتدا عملکرد الگوریتم تشخیص و شمارش کاکل‌های گیاه ذرت MYOLOv5s با الگوریتم YOLOv5s مقایسه می‌شود و پس از اثبات بهبود عملکرد الگوریتم در اثر اعمال تغییرات این الگوریتم با سایر الگوریتم‌های طراحی‌شده در این زمینه مقایسه و عملکرد هر یک بررسی می‌شود. همچنین تاثیر امتیاز بر عملکرد الگوریتم‌های مبتنی بر تشخیص و شمارش کاکل‌های گیاه ذرت نیز مورد ارزیابی قرار می‌گیرد.

مقایسه مدل‌های اصلی و بهبودیافته YOLOv5s

در جدول ۱ مقایسه‌ای بین نتایج حاصل از تشخیص کاکل‌های ذرت در مرحله آزمودن الگوریتم‌های YOLOv5s و MYOLOv5s به ازای ۸۰ درصد داده آموزش، ۱۰ درصد داده اعتبارسنجی و ۱۰ درصد داده آزمون انجام گرفته است. چنانچه که از جدول پیداست، با اعمال اصلاحات در معماری شبکه، ۲/۸۰ درصد بهبود در دقت تشخیص کاکل‌های ذرت حاصل می‌شود که بیانگر افزایش تعداد کاکل‌های ذرت به‌درستی تشخیص داده شده توسط روش پیشنهادی و موثر بودن نوآوری‌های درنظر گرفته شده است.

جدول ۱– مقایسه دقت میانگین روش‌های YOLOv5s و MYOLOv5s

Table 1- Comparison between the AP values of YOLOv5s and MYOLOv5s algorithms

الگوریتم **Algorithm**	دقت میانگین **AP (%)**
YOLOv5s	92.50
MYOLOv5s	95.30

مقایسه با سایر روش‌های تشخیص و شمارش شی

1- Root Mean Square Error
2- Frame Per Second

بررسی دقت تشخیص و شمارش روش‌ها

عملکرد الگوریتم MYOLOv5s در تشخیص و شمارش کاکل‌های ذرت تصاویر آزمون با یک الگوریتم مبتنی بر شمارش TasselNetv2+ (Lu & Cao, 2020) و دو الگوریتم مبتنی بر تشخیص Faster R-CNN (Liu *et al.*, 2020) و RetinaNet (Zou *et al.*, 2020) مقایسه شده است. TasselNetv2+ و RetinaNet در کتابخانه پایتورچ پیاده‌سازی شده‌اند و الگوریتم Faster R-CNN مبتنی بر کتابخانه تنسورفلو[3] (Dillon *et al.*, 2017) می‌باشد. همچنین، برای مقایسه روش‌ها در شرایط برابر، از قرینه‌سازی[4] افقی ۵۰ درصد از داده‌ها جهت داده‌افزایی[5] استفاده می‌شود و بهترین عملکرد هر الگوریتم به‌ازای تعداد تکرارها و اندازه دسته بهینه در نظر گرفته شده است. داده‌های آموزش، اعتبارسنجی و آزمون استفاده شده نیز برای تمامی الگوریتم‌ها یکسان هستند. جزییات آموزش روش‌ها در جدول ۲ آورده شده است.

همان‌گونه که در شکل ۴ نشان داده شده است مقادیر اندازه دسته و تکرار برای مدل MYOLOv5s به‌صورت بهینه به ازای بهترین عملکرد الگوریتم به‌دست آمده‌اند. در این شکل مقادیر دقت‌های میانگین به‌دست آمده برای الگوریتم MYOLOv5s به‌ازای اندازه دسته‌های مختلف ۴، ۸ و ۱۶ در ۵۰، ۱۰۰، ۱۵۰، ۲۰۰ و ۲۵۰ تکرار نشان داده شده است. همان‌گونه که از شکل پیداست عملکرد بهینه شبکه به ازای ۲۰۰ تکرار و اندازه دسته ۱۶ به‌دست می‌آید و پس از آن بیش‌برازش رخ می‌دهد. لازم به ذکر است که مقادیر اندازه دسته بیشتر از ۱۶ به علت عدم ظرفیت سیستم پردازشی (محدودیت RAM مربوط به کارت گرافیک مورد استفاده) بررسی نشده‌اند. برای سایر الگوریتم‌ها نیز پارامترهای بهینه نیز با توجه به مقادیر گزارش‌شده در مقالات مربوطه، پایگاه داده موجود و محدودیت توان سیستم پردازشی انتخاب شده‌اند.

نمودار همگرایی دقت میانگین الگوریتم MYOLOv5s نیز در شکل ۵ نشان داده شده است. با توجه به این منحنی انتخاب ۲۰۰ تکرار جهت آموزش شبکه مناسب می‌باشد.

نتایج حاصل از دقت میانگین و R^2 بر روی داده‌های اعتبارسنجی و زمان آموزش مورد نیاز در جدول ۳ آورده شده است که نشان می‌دهد روش پیشنهادی ویژگی‌های کاکل‌های ذرت را بیشتر یاد گرفته و با دقت بالاتری آن‌ها را تشخیص می‌دهد. همچنین زمان آموزش MYOLOv5s در مقایسه با دیگر روش‌های بررسی‌شده کمتر است.

نتایج کمی و بصری تشخیص و شمارش کاکل‌های ذرت توسط

3- Tensorflow
4- Flip
5- Data Augmentation

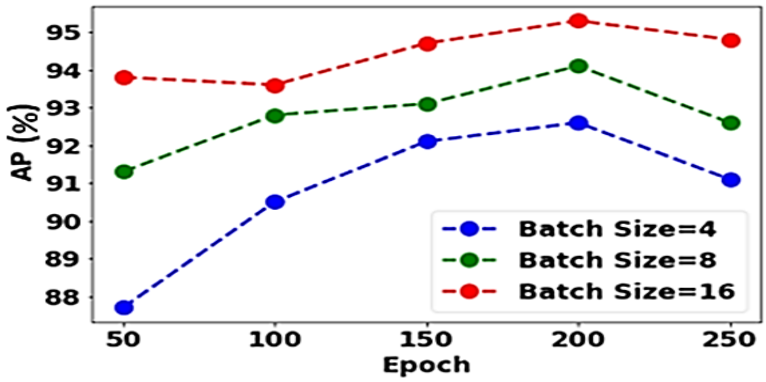
الگوریتم‌های نامبرده در زمان آزمودن نیز در جدول ۴ و شکل ۶ آمده است.

از مقایسه نتایج حاصل از جداول ۳ و ۴ و نزدیک بودن مقادیر به‌دست‌آمده آموزش و آزمودن شبکه‌ها می‌توان دریافت که مقادیر پارامترهای آموزش الگوریتم‌ها به خوبی انتخاب گردیده و بیش‌برازش رخ نداده است. همان‌گونه که از جدول ۴ پیداست الگوریتم MYOLOv5s قادر می‌باشد که مقادیر دقت، یادآوری و دقت میانگین معادل با ۹۳/۵۸، ۹۴/۹۷ و ۹۵/۳۰ درصد را به‌ترتیب به‌دست آورد.

جدول ۲- تنظیم پارامترهای مجموعه داده HIPS توسط الگوریتم‌های مختلف

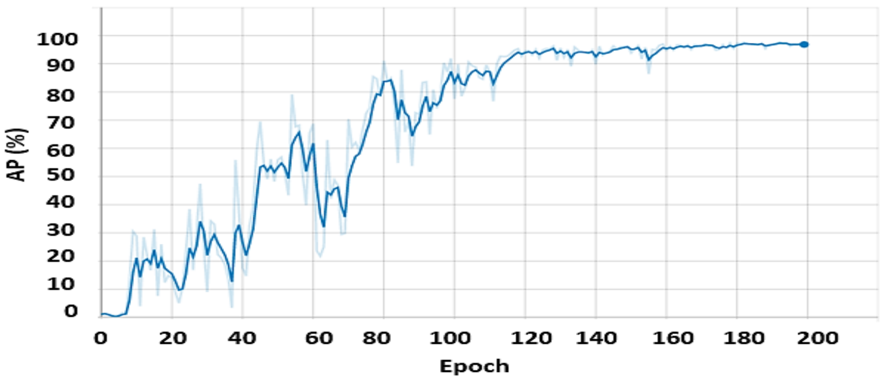
Table 2- Parameter setting of HIPS dataset by different methods

الگوریتم Algorithm	تعداد تصاویر آموزش NO. Train	تعداد تصاویر اعتبارسنجی NO. Validation	تعداد تصاویر آزمون NO. Test	اندازه دسته Batch Size	تعداد تکرار Epoch	روش بهینه‌سازی Optimization method
TasselNetv2+	270 (mixed with validation data)	–	30	16	300	SGD
Faster R-CNN	240	30	30	4	833	SGD
RetinaNet	240	30	30	4	500	Adam
MYOLOv5s	240	30	30	16	200	Adam



شکل ٤- تعیین مقدار بهینه پارامترهای اندازه دسته و تکرار MYOLOv5s

Fig.4. MYOLOv5s determination of the optimal batch-size and epoch values



شکل ۵- منحنی همگرایی دقت میانگین الگوریتم MYOLOv5s

Fig.5. AP convergence plot of MYOLOv5s algorithm

جدول ۳- مقایسه دقت میانگین و R^2 به‌دست‌آمده توسط الگوریتم‌های مختلف در مرحله آموزش

Table 3- Comparison among the AP and R^2 values obtained by different algorithms in training step

الگوریتم Algorithm	دقت میانگین AP (%)	R^2 (%)	تعداد فریم بر ثانیه FPS
TasselNetv2+	–	80.23	61
Faster R-CNN	80.83	79.71	13
RetinaNet	81.26	99.44	29
MYOLOv5s	98.69	99.45	84

جدول ٤- مقایسه عملکرد و سرعت پردازش روش‌های مختلف بر روی تصاویر آزمون مجموعه داده HIPS

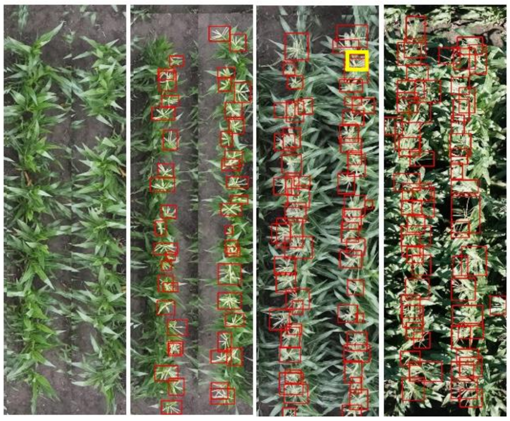
Table 4- Comparison of performance and inference time of different methods on HIPS test images

الگوریتم Algorithm	دقت Precision (%)	یادآوری Recall (%)	دقت میانگین AP (%)	خطای جذر میانگین مربعات RMSE	R^2 (%)	تعداد فریم بر ثانیه FPS
TasselNetv2+	–	–	–	10.32	77.86	61
Faster R-CNN	91.35	82.85	76.99	8.58	86.83	13
RetinaNet	92.49	88.04	77.66	6.01	95.53	29
MYOLOv5s	93.58	94.97	95.30	1.9	99.28	84

این مقادیر برای الگوریتم Faster R-CNN به‌ترتیب ۹۱/۳۵، ۸۲/۸۵، ۷۶/۹۹ درصد و برای الگوریتم RetinaNet به‌ترتیب ۹۲/۴۹، ۸۸/۰۴ و ۷۷/۶۶ درصد گزارش شده است. مقدار بالای دقت الگوریتم MYOLOv5s نسبت به سایر الگوریتم‌ها حاکی از عملکرد بهتر آن در تشخیص درست تعداد بیشتری از کاکل‌های ذرت نسبت دو الگوریتم دیگر است. الگوریتم Faster R-CNN نسبت به الگوریتم‌های MYOLOv5s و RetinaNet مقدار یادآوری کمتری را کسب کرده که نمایانگر این است در این الگوریتم بسیاری از کاکل‌های ذرت تشخیص داده نشده‌اند. این مقدار برای RetinaNet نیز در مقایسه با MYOLOv5s، ۶/۹۳ درصد کمتر است. لذا، با مقایسه این مقادیر می‌توان دریافت که تعداد زیادی از کاکل‌های گیاه ذرت با توجه به شکل خاص، متفاوت و هم‌پوشانی بالایی که دارند به آسانی توسط الگوریتم‌های Faster-RCNN و RetinaNet قابل تشخیص نیستند و تنها کاکل‌های واضح توسط این الگوریتم‌ها شناسایی می‌شوند حال آن‌که الگوریتم MYOLOv5s قادر خواهد بود کاکل‌ها با اندازه‌ها و شکل‌های متفاوت و در نواحی با هم‌پوشانی بالا را شناسایی کند لذا تعداد بیشتری از کاکل‌های ذرت توسط این الگوریتم قابل شناسایی‌اند و مقدار یادآوری بالاتری را کسب می‌کند. با بررسی نتایج حاصل از دقت میانگین به‌عنوان معیاری استاندارد که مقادیر دقت و یادآوری را برای سنجش الگوریتم توأماً در نظر می‌گیرد، مجدداً اثبات می‌گردد که در بین الگوریتم‌های مورد بررسی Faster-RCNN ضعیف‌ترین عملکرد را داراست و الگوریتم MYOLOv5s با کمترین مقدار تشخیص‌های اشتباه و یا انجام نگرفته بهترین عملکرد را در کوتاه‌ترین زمان در مقایسه با سایر روش‌ها دارد.

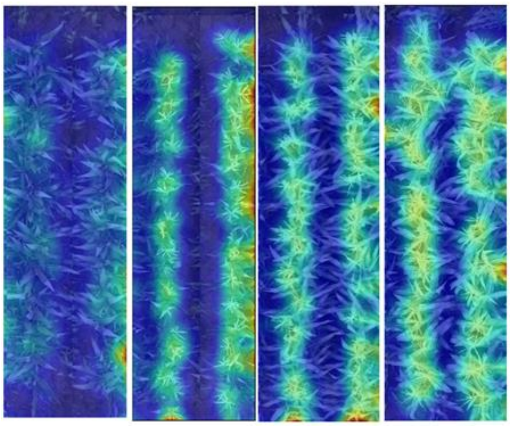
معیارهای خطای جذر میانگین مربعات و R^2 نیز برای ارزیابی دقت الگوریتم‌ها در شمارش کاکل‌های ذرت با یکدیگر مقایسه شده‌اند. این مقادیر برای الگوریتم MYOLOv5s به‌ترتیب ۱/۹ و ۹۹/۲۸ درصد، برای الگوریتم‌های مبتنی بر تشخیص Faster-RCNN و RetinaNet به‌ترتیب ۸/۵۸، ۸۶/۸۳ و ۶/۰۱، ۹۵/۵۳ درصد و برای الگوریتم مبتنی بر شمارش +TasselNetv2 به‌ترتیب ۱۰/۳۲ و ۷۷/۸۶ درصد به‌دست آمده‌اند. مقدار کم خطای جذر میانگین مربعات حاصل آمده توسط الگوریتم MYOLOv5s، نزدیک بودن تعداد کاکل‌های ذرت تخمین زده توسط این الگوریتم به تعداد حقیقی آن‌ها را نشان می‌دهد. حال آن‌که این مقدار برای الگوریتم‌های Faster-RCNN، RetinaNet و +TasselNetv2 بیشتر گزارش شده است که به معنای عدم عملکرد دقیق الگوریتم‌های نامبرده در شمارش کاکل‌های ذرت به‌ویژه در نواحی با هم‌پوشانی بالا است. مقدار R^2 نیز برای الگوریتم MYOLOv5s بیشتر از سه الگوریتم بررسی شده دیگر به‌دست آمده است که خود به معنای همبستگی بیشتر تعداد کاکل‌های ذرت تخمین زده شده و مقادیر حقیقی آن‌ها در این الگوریتم می‌باشد.

با بررسی نتایج بصری شکل ۶ عملکرد روش‌های مختلف بهتر نمایان می‌شود.



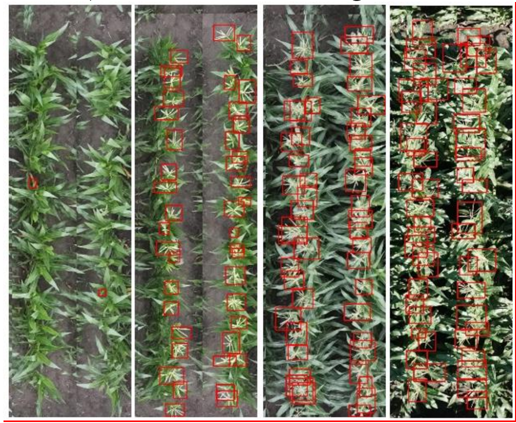
الف) کادرهای محصورکننده حقیقی

a) Ground Truth Bounding Boxes



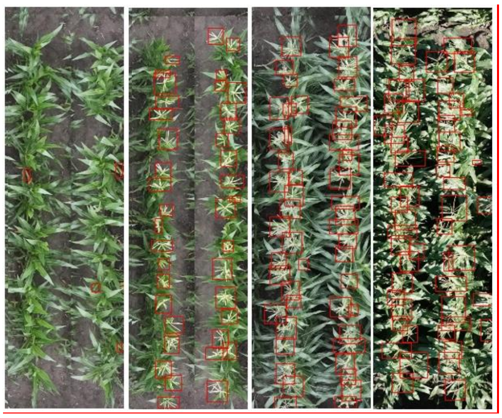
ب) +TasselNetv2

b) TasselNetv2+



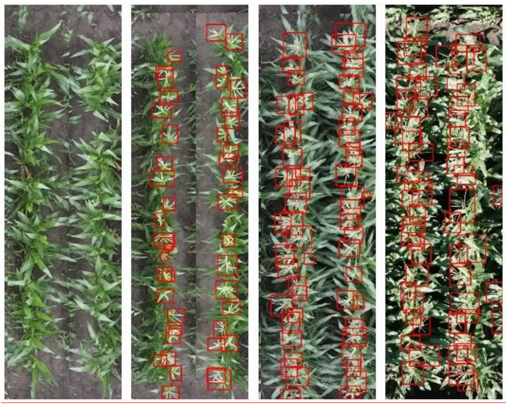
ج) Faster R-CNN

c) Faster R-CNN



د) RetinaNet

d) RetinaNet



ه) MYOLOv5s

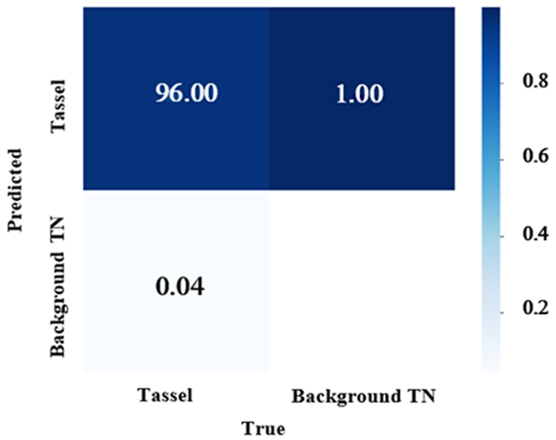
e) MYOLOv5s

شکل ٦- نتایج بصری کادرهای محصورکننده حقیقی و نتایج به‌دست‌آمده توسط الگوریتم‌های +TasselNetv2، Faster R-CNN، RetinaNet و MYOLOv5s

Fig.6. Visual results of ground truth bounding boxes and the obtained results by TasselNetv2+, Faster R-CNN, RetinaNet, and MYOLOv5s algorithms

این شکل متشکل از ۴ تصویر از مجموعه داده آزمون است که دو تصویر اول از سمت راست کاکل‌های ذرت هیبرید را نشان می‌دهند و دو تصویر بعدی مربوط به کاکل‌های ذرت اینبرد هستند. همان‌گونه که در تصاویر قابل رویت است دو نوع ذرت نامبرده شکل و میزان رشد متفاوتی دارند. با بررسی شکل می‌توان دریافت که الگوریتم‌هایFaster-RCNN و RetinaNet کاکل‌های ذرت را در نواحی متراکم نمی‌توانند به‌خوبی تشخیص دهند. کاکل مربوط به ذرت هیبرید مشخص شده با رنگ زرد در تصویر کادرهای محصورکننده حقیقی از این دست است که از بین مدل‌های مبتنی بر تشخیص تنها مدل MYOLOv5s آن را شناسایی کرده است. همچنین تعداد تشخیص‌های اشتباه الگوریتم‌های Faster R-CNN و RetinaNet نیز نسبت به MYOLOv5s بیشتر است. این امر را می‌توان در تصویر دوم مربوط به ذرت اینبرد مشاهده کرد که علی‌رغم عدم وجود کاکل ذرت در آن الگوریتم‌های Faster-RCNN و RetinaNet تعدادی را به اشتباه تشخیص داده‌اند. به‌علاوه، کادرهای محصورکننده تخمین‌زده‌شده توسط الگوریتم MYOLOv5s از منظر موقعیت قرارگیری و اندازه شباهت بیشتری به کادرهای محصورکننده حقیقی نسبت به سایر روش‌ها دارند لذا این الگوریتم عملکرد بهتری دارد. این در حالی است که در روش مبتنی بر شمارش TasselNetv2+، به علت گزارش تنها تخمینی از تعداد کاکل‌های ذرت، نمی‌توان در مورد موقعیت و تشخیص صحیح آن‌ها توسط الگوریتم اظهار نظر نمود. بنابراین تنها می‌توان ادعا کرد که با توجه به مقدار پایین R^2 و بالای RMSE به‌دست آمده توسط TasselNetv2+ اختلاف معناداری بین مقادیر حقیقی و مقادیر پیش‌بینی‌شده توسط آن وجود دارد که خود نمایانگر تعداد زیاد تشخیص‌های اشتباه صورت‌گرفته توسطTasselNetv2+ می‌باشد.

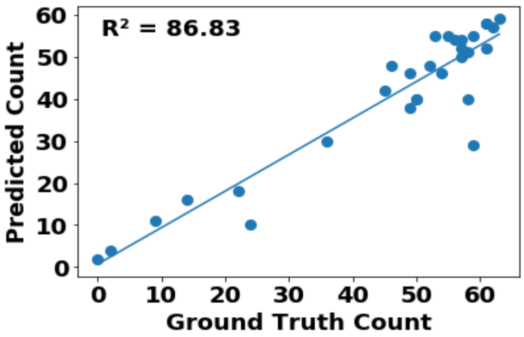
ماتریس اغتشاش الگوریتم MYOLOv5s در شکل ۷ نشان داده شده است. با توجه به شکل، ۹۶ درصد از کاکل‌های ذرت درست تشخیص داده شده‌اند و ۴ درصد دیگر شناسایی نشده‌اند که با توجه به پیچیدگی پایگاه داده دقت قابل‌توجهی برای مدل MYOLOv5s محسوب می‌گردد.



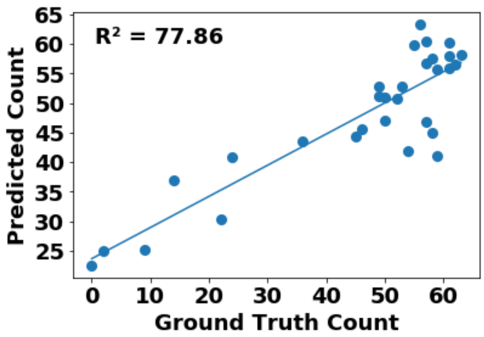
شکل ۷- ماتریس اغتشاش مربوط به داده‌های آزمون الگوریتم MYOLOv5s

Fig.7. MYOLOv5s confusion matrix of test images

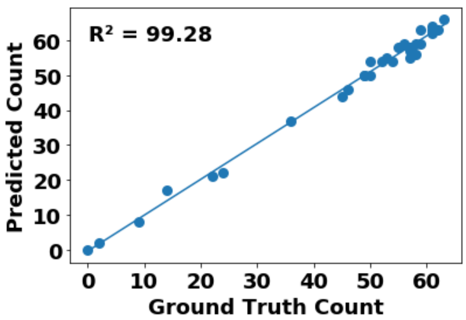
به‌منظور ارزیابی دقیق‌تر عملکرد الگوریتم‌ها در شمارش، از منحنی رگرسیون خطی استفاده می‌شود. شکل ۸ این منحنی‌ها را برای الگوریتم‌های به‌کاربرده در این مقاله نشان می‌دهد. همان‌گونه که از منحنی‌ها پیداست، نقاط در الگوریتم MYOLOv5s متراکم‌ترند. لذا، با توجه به نزدیک‌تر بودن تعداد کاکل‌های ذرت تخمین‌زده‌شده و حقیقی توسط این الگوریتم می‌توان گفت الگوریتم MYOLOv5s نسبت به سایر الگوریتم‌ها قادر به تخمین دقیق‌تر تعداد کاکل‌های ذرت می‌باشد. در این بین منحنی مربوط به TasselNetv2+ پراکنده‌ترین نقاط و در نتیجه ضعیف‌ترین عملکرد را داراست.



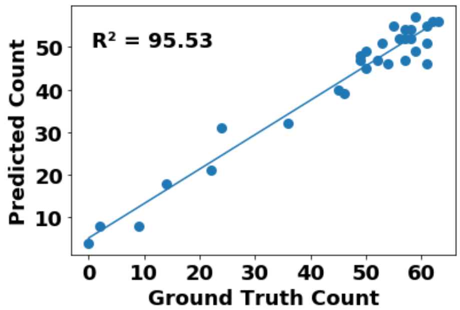
ب) Faster R-CNN
b) Faster R-CNN



الف) TasselNetv2+
a) TasselNetv2+



د) MYOLOv5s
d) MYOLOv5s



ج) RetinaNet
c) RetinaNet

شکل ۸– منحنی‌های رگرسیون خطی الگوریتم‌های TasselNetv2+، Faster R-CNN، RetinaNet و MYOLOv5s

Fig.8. Linear regression plots of TasselNetv2+, Faster R-CNN, RetinaNet, and MYOLOv5s algorithms

جدول ۵- مقادیر امتیاز (٪) F1 تصاویر آزمون مجموعه داده HIPS برای الگوریتم‌های مبتنی بر تشخیص Faster R-CNN، RetinaNet و MYOLOv5s

Table 5- F1 (%) score values of HIPS test images obtained by Faster R-CNN, RetinaNet and MYOLOv5s algorithms

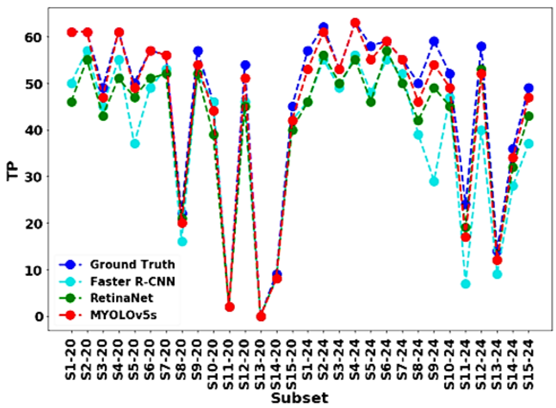
تصویر آزمون Test subset	الگوریتم Algorithm			تصویر آزمون Test subset	الگوریتم Algorithm		
	Faster R-CNN	**RetinaNet**	**MYOLOv5s**		**Faster R-CNN**	**RetinaNet**	**MYOLOv5s**
S1–20	88.53	85.97	98.37	**S1–24**	85.98	88.47	94.67
S2–20	95.79	94.85	97.59	**S2–24**	92.44	94.9	97.59
S3–20	94.71	88.69	94.94	**S3–24**	90.77	96.11	98.17
S4–20	92.44	91.07	99.19	**S4–24**	91.8	92.43	97.7
S5–20	82.22	94.94	94.21	**S5–24**	88.09	83.65	93.99
S6–20	89.91	93.15	99.14	**S6–24**	96.48	98.27	96.75
S7–20	96.32	96.32	97.38	**S7–27**	94.5	90.9	97.33
S8–20	79.99	97.7	93	**S8–24**	86.67	88.41	92
S9–20	93.68	93.68	94.7	**S9–24**	65.95	90.77	91.5
S10–20	97.85	91.77	95.7	**S10–24**	94	90.87	92.42
S11–20	66.67	65	100	**S11–24**	41.21	69.11	73.91
S12–20	92.01	89.97	94.4	**S12–24**	81.66	94.63	91.27
S13–20	–	–	–	**S13–24**	60.03	75.02	77.42
S14–20	79.99	94.12	94.12	**S14–24**	84.85	94.12	93.13
S15–20	96.53	94.12	94.39	**S15–24**	89.93	89.61	94.94
کل Total	89.045	90.81	96.22	**کل Total**	82.96	89.15	92.19

بررسی اثرات زمان کاشت در عملکرد روش‌های مبتنی بر تشخیص

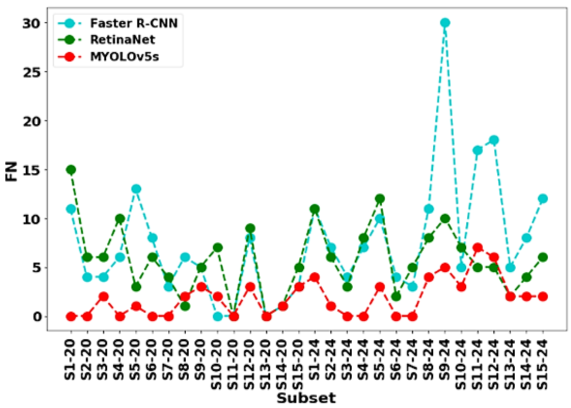
در جدول ۵ مقایسه‌ای بین روش‌های مبتنی بر تشخیص در شناسایی دقیق موقعیت کاکل‌های ذرت تصاویر آزمون برای دو نوع ذرت هیبرید و اینبرد و در دو تاریخ مختلف بیستم و بیست و چهارم جولای آورده شده است.

تصاویر مشخص‌شده با عنوان S1-S7 کاکل‌های ذرت هیبرید را نشان می‌هند و تصاویر S8-S15 مربوط به ذرت از نوع اینبرد هستند. عددهای ۲۰ و ۲۴ نیز اشاره به تاریخ تصویربرداری در این نام‌گذاری می‌کنند. جهت ارزیابی عملکرد الگوریتم‌های مختلف و تاثیر زمان بر روی دقت تشخیص الگوریتم‌ها از معیار امتیاز F1 در این جدول استفاده شده است. نتایج به‌دست آمده نشان می‌دهند که در تاریخ بیست و چهارم با افزایش رشد کاکل‌های گیاه ذرت به علت تراکم بالای آن‌ها تشخیص توسط الگوریتم‌های مختلف دشوارتر می‌گردد. علاوه بر آن به علت نور زیاد محیط در زمان تصویربرداری و خشک بودن خاک تعداد تشخیص‌های اشتباه برای این تاریخ نسبت به تاریخ بیستم نیز بیشتر است. در بین الگوریتم‌های مورد بررسی MYOLOv5s کاکل‌های ذرت اکثر تصاویر آزمون را بهتر از الگوریتم‌های Faster R-CNN و RetinaNet تشخیص می‌دهد و بیشترین تعداد تشخیص‌های درست را داراست. این الگوریتم قادر است که کاکل‌های ذرت را در تاریخ ۲۰ جولای با اختلاف به‌ترتیب ۷/۱۸ و ۵/۴۱ درصد نسبت به الگوریتم‌های Faster R-CNN و RetinaNet و در تاریخ ۲۴ جولای با اختلاف ۹/۲۳ و ۳/۰۴ درصد بهتر از دو الگوریتم دیگر شناسایی کند.

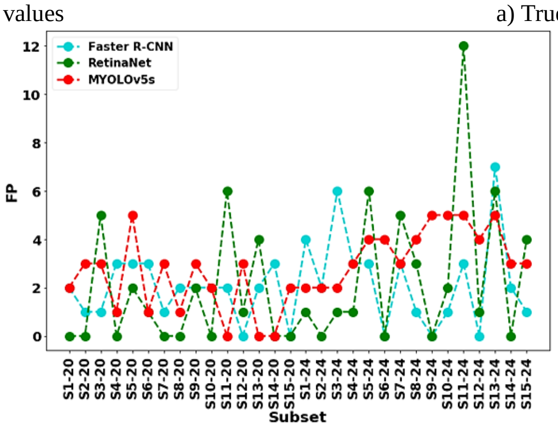
شکل ۹ نیز منحنی‌های تعداد TP، تعداد FP و تعداد FN را برای سه الگوریتم مبتنی بر تشخیص مورد بحث در دو زمان نشان می‌دهد.



الف) تعداد مثبت‌های صحیح
a) True positive values



ب) تعداد منفی‌های کاذب
b) False negative values



ج) تعداد مثبت‌های کاذب
c) False positive values

شکل ۹– تعداد تشخیص‌های مثبت صحیح (TP)، منفی کاذب (FN) و مثبت کاذب (FP) به‌دست‌آمده توسط الگوریتم‌های مبتنی بر تشخیص Faster R-CNN، RetinaNet و MYOLOv5s

Fig.9. TP, FN, and FP values obtained by detection-based Faster R-CNN, RetinaNet, and MYOLOv5s algorithms

همان‌گونه که در این شکل پیداست، در دو تاریخ مورد بررسی الگوریتم MYOLOv5s بیشترین تعداد تشخیص‌های درست را داراست. با توجه به شکل این الگوریتم در تشخیص‌های صحیح بسیار مشابه با منحنی تعداد حقیقی کاکل‌های ذرت (Ground truth) عمل کرده است. به‌علاوه، تعداد تشخیص‌های نادرست آن نسبت به کل تشخیص‌های درست انجام‌گرفته توسط این الگوریتم تنها ۶/۴۲ درصد است در حالی‌که این نسب برای روش‌های Faster R-CNN و RetinaNet به‌ترتیب ۸/۶۵ و ۷/۵۱ درصد می‌باشد. در این شکل تعداد تشخیص‌های صحیح انجام‌گرفته توسط Faster R-CNN کمتر از سایر روش‌هاست. همچنین، هر سه نمودار کلیه الگوریتم‌ها برای داده‌های تاریخ بیست و چهارم نسبت به بیستم تغییرات بیشتری را نشان می‌دهند که خود ثابت‌کننده عملکرد ضعیف‌تر الگوریتم‌ها در این تاریخ به علت چالش‌های موجود می‌باشد. استفاده از این شکل تحلیل بر اساس نوع ذرت را نیز ملموس‌تر می‌سازد و بر این اساس می‌توان گفت که در تصاویر مربوط به ذرت هیبرید که ذرت‌ها در آن رشد بیشتری داشته‌اند و متراکم‌تر واقع شده‌اند تعداد تشخیص‌های صحیح انجام‌گرفته توسط MYOLOv5s بیشتر از سایر الگوریتم‌هاست. Faster R-CNN نیز در تصاویر مربوط به ذرت اینبرد تاریخ بیست و چهارم جولای به‌صورت محسوسی ضعیف‌تر عمل می‌کند که نشان‌دهنده عدم توانایی این الگوریتم در شناسایی کاکل‌های ریز است.

همچنین، با توجه به جدول ۵، می‌توان تصویر S11-24 را به‌عنوان چالش‌برانگیزترین داده با پایین‌ترین دقت تشخیص از بین تصاویر مجموعه داده HIPS دانست. این اتفاق به علت تفاوت در شدت نور و رنگ داده‌های آموزش و آزمون این تصویر رخ داده است. شکل ۱۰ نمونه‌ای از داده آموزش و ارزیابی متعلق به این تصویر را نشان می‌دهد.



الف) تصویر آموزش
a) Train image



ب) تصویر آزمون
b) Test image

شکل ۱۰- نمونه‌ای از تصاویر آموزش و آزمون داده S11 تاریخ ۲۴ جولای (S11-24)

Fig.10. Train and test images S11 on 24th July (S11–24)

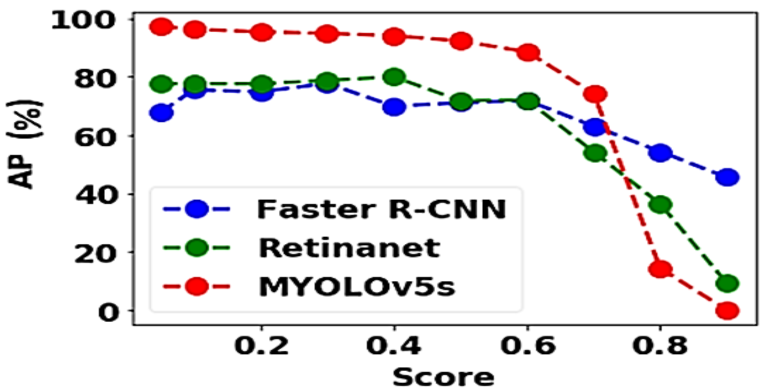
بررسی تاثیر امتیاز

امتیاز تشخیص شی نقش به‌سزایی در عملکرد مدل‌های مبتنی بر تشخیص ایفا می‌کند. برای مطالعه اثر این پارامتر در مدل‌های مختلف، تغییرات مقدار دقت میانگین (AP) به ازای مقادیر مختلف امتیاز در منحنی شکل ۱۱ آمده است. همان‌گونه که از شکل پیداست امتیاز بر روی عملکرد مدل‌های تک مرحله‌ای نظیر MYOLOv5s و RetinaNet تاثیر بیشتری دارد. مقدار بهینه این پارامتر برای الگوریتم‌های Faster R-CNN، RetinaNet و MYOLOv5s به‌ترتیب ۰/۲۶، ۰/۲۸ و ۰/۲ به‌دست آمد.

نتیجه‌گیری

در این مقاله، جهت نظارت مستمر بر روند رشد گیاه ذرت با هدف برداشت حداکثری این گیاه به‌عنوان یکی از مهم‌ترین غلات، الگوریتمی جهت تشخیص و شمارش خودکار کاکل‌های ذرت بر پایه الگوریتم عمیق YOLOv5s پیشنهاد شده است. جهت آموزش و

ارزیابی این الگوریتم از مجموعه داده HIPS دانشگاه پردو، ایندیانا، آمریکا مربوط به سال ۲۰۲۰ متشکل از تصاویر پهپادی برداشت شده از کاکل‌های گیاه ذرت در دو تاریخ مختلف با میزان رشد متفاوت استفاده شده است.



شکل ۱۱– منحنی دقت میانگین بر اساس امتیاز برای الگوریتم‌های مبتنی بر تشخیص Faster R-CNN، RetinaNet و MYOLOv5s

Fig.11. AP versus score plot for detection-based methods TasselNetv2+, Faster R-CNN, RetinaNet, and MYOLOv5s

با توجه به کوچک بودن مجموعه داده و ویژگی‌های کاکل‌های ذرت در تصاویر (شکل، اندازه، رنگ، بافت و غیره) به‌منظور هرچه بهتر عمل کردن الگوریتم اصلاحات اساسی در معماری الگوریتم اصلی صورت‌گرفته و الگوریتم بهبودیافته مختص شناسایی کاکل‌های ذرت با نام MYOLOv5s ارائه شده است. بهبودهای صورت‌گرفته شامل استفاده از ساختار BottleneckCSP در گردن شبکه و تابع فعال‌سازی Hardswish جهت یادگیری تعداد بیشتری ویژگی توسط شبکه و بهبود عملکرد آن و استفاده از لایه‌های حذف تصادفی مکانی در سر شبکه جهت جلوگیری از بیش‌برازش و افزایش دقت تشخیص به‌ویژه در زمان کوچک بودن مجموعه داده می‌شود. عملکرد این شبکه در تشخیص و شمارش کاکل‌های ذرت با روش‌های معتبر معرفی‌شده در سال‌های اخیر Faster R-CNN، RetinaNet و TasselNetv2+ مقایسه گردیده و نتایج عددی به‌دست‌آمده نشان دادند که الگوریتم MYOLOv5s بیشترین تعداد تشخیص صحیح کاکل‌های ذرت با بالاترین دقت میانگین و کم‌ترین زمان پردازش را دارا می‌باشد. تصاویر حاصل از کاکل‌های ذرت تشخیص داده شده توسط چند روش مورد بررسی نیز اثبات کردند که این الگوریتم بیش از سایر روش‌ها قادر به تشخیص کاکل‌های ذرت با اندازه‌های متفاوت است و در نواحی متراکم نیز عملکرد خوبی دارد.

در کارهای آینده از ایده یادگیری انتقالی جهت شمارش گیاهان مشابه مانند سویا و پنبه استفاده خواهد شد. همچنین با به‌کارگیری شبکه‌های عمیق متفاوت و بهبود هرچه بیشتر آن‌ها و استفاده از روش‌های داده افزایی مبتنی بر یادگیری عمیق نظیر تولید داده مصنوعی در جهت عملکرد هرچه بهتر روش‌های یادگیری عمیق در حوزه کشاورزی اقدام خواهد شد. به‌علاوه، با تصویربرداری از گیاهان با سنجنده‌های دیگر نظیر دوربین چند طیفی روش‌هایی برای تعیین خودکار آفات و بیماری‌های گیاهان تعبیه خواهد شد.

سپاسگزاری

این پژوهش مربوط به مرکز تحقیقات و آموزش زراعت (ACRE) دانشگاه پردو، ایندیانا، آمریکا است. بر خود لازم می‌دانیم از شرکت دانش بنیان فنی مهندسی داده کاوی ویرا ژرف کویر که تصاویر پایگاه داده را تهیه نموده و در به ثمر رسیدن این پژوهش نقش به‌سزایی داشت تقدیر و تشکر نماییم.

References

1. Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). *Understanding of a convolutional neural network*. Pages 1-6. 2017 International Conference on Engineering and Technology (ICET): IEEE. https://doi.org/10.1109/ICEngTechnol.2017.8308186
2. Alzadjali, A., Alali, M. H., Sivakumar, A. N. V., Deogun, J. S., Scott, S., Schnable, J. C., & Shi, Y. (2021). Maize Tassel Detection from UAV Imagery Using Deep Learning. Frontiers in Robotics and AI. https://doi.org/10.3389/frobt.2021.600410

3. Bisong, E. (2019). Building machine learning and deep learning models on Google Cloud Platform. Springer.
4. Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934. https://doi.org/10.48550/arXiv.2004.10934
5. Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *Computer Science*, *7*, e623. https://doi.org/10.7717/peerj-cs.623
6. Dillon, J. V., Langmore, I., Tran, D., Brevdo, E., Vasudevan, S., Moore, D., Patton, B., Alemi, A., Hoffman, M., & Saurous, R. A. (2017). Tensorflow distributions. arXiv preprint arXiv: 1711.10604. https://doi.org/10.48550/arXiv.1711.10604
7. Elfwing, S., Uchibe, E., & Doya, K. (2018). Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Networks*, *107*, 3-11. https://doi.org/10.1016/j.neunet.2017.12.012
8. Farhadi, A., & Redmon, J. (2018). Yolov3: An incremental improvement. Pages 1804-2767. Computer Vision and Pattern Recognition: Springer Berlin/Heidelberg, Germany. https://doi.org/10.48550/arXiv.1804.02767
9. Ghosal, S., Zheng, B., Chapman, S. C., Potgieter, A. B., Jordan, D. R., Wang, X., Singh, A. K., Singh, A., Hirafuji, M., & Ninomiya, S. (2019). A weakly supervised deep learning framework for sorghum head detection and counting. *Plant Phenomics* 2019. https://doi.org/10.34133/2019/1525874
10. Gómez-Flores, W., Garza-Saldaña, J. J., & Varela-Fuentes, S. E. (2019). Detection of huanglongbing disease based on intensity-invariant texture analysis of images in the visible spectrum. *Computers and Electronics in Agriculture*, *162*, 825-835. https://doi.org/10.1016/j.compag.2019.05.032
11. Habib, A. F., Kim, E. M., & Kim, C. J. (2007). New methodologies for true orthophoto generation. *Photogrammetric Engineering & Remote Sensing*, *73*, 25-36.
12. Hawkins, D. M. (2004). The problem of overfitting. *Journal of Chemical Information and Computer Sciences*, *44*, 1-12. https://doi.org/10.1021/ci0342472
13. He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *37*, 1904-1916. https://doi.org/10.1109/TPAMI.2015.2389824
14. Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., & Vasudevan, V. (2019). *Searching for mobilenetv3*. Pages 1314-1324. Proceedings of the IEEE/CVF International Conference on Computer Vision.
15. Jocher, G., et al. (2020). Yolov5. https://github.com/ultralytics/yolov5
16. Lempitsky, V., & Zisserman, A. (2010). Learning to count objects in images. *Advances in Neural Information Processing Systems*, *23*, 1324-1332.
17. Leung, H., & Haykin, S. (1991). The complex backpropagation algorithm. *IEEE Transactions on Signal Processing*, *39*, 2101-2104. https://doi.org/10.1109/78.134446
18. Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). *Feature pyramid networks for object detection*. Pages 2117-2125. Proceedings of the IEEE conference on computer vision and pattern recognition.
19. Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. (2018). *Path aggregation network for instance segmentation*. Pages 8759-8768. Proceedings of the IEEE conference on computer vision and pattern recognition.
20. Liu, Y., Cen, C., Che, Y., Ke, R., Ma, Y., & Ma, Y. (2020). Detection of maize tassels from UAV RGB imagery with faster R-CNN. *Remote Sensing*, *12*, 338. https://doi.org/10.3390/rs12020338
21. Long, X., Deng, K., Wang, G., Zhang, Y., Dang, Q., Gao, Y., Shen, H., Ren, J., Han, S., & Ding, E. (2020). PP-YOLO: An effective and efficient implementation of object detector. arXiv preprint arXiv:2007.12099. https://doi.org/10.48550/arXiv.2007.12099
22. Lu, H., & Cao, Z. (2020). Tasselnetv2+: A fast implementation for high-throughput plant counting from high-resolution RGB imagery. *Frontiers in Plant Science*, *11*, 1929. https://doi.org/10.3389/fpls.2020.541960
23. Lu, H., Cao, Z., Xiao, Y., Zhuang, B., & Shen, C. (2017). TasselNet: counting maize tassels in the wild via local counts regression network. *Plant Methods*, *13*, 1-17. https://doi.org/10.1186/s13007-017-0224-0
24. Ongsulee, P. (2017). *Artificial intelligence, machine learning and deep learning*. Pages 1-6. 2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE): IEEE.
25. Parihar, C., Jat, S., Singh, A., Kumar, R. S., Hooda, K., GK, C., & Singh, D. (2011). Maize production technologies in India.
26. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., & Antiga, L. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, *32*, 8026-8037.
27. Pourreza, A., Lee, W. S., Etxeberria, E., & Banerjee, A. (2015). An evaluation of a vision-based sensor performance in Huanglongbing disease identification. *Biosystems Engineering*, *130*, 13-22. https://doi.org/10.1016/j.biosystemseng.2014.11.013

28. Quan, L., Feng, H., Lv, Y., Wang, Q., Zhang, C., Liu, J., & Yuan, Z. (2019). Maize seedling detection under different growth stages and complex field environments based on an improved Faster R–CNN. *Biosystems Engineering*, *184*, 1-23. https://doi.org/10.1016/j.biosystemseng.2019.05.002
29. Redmon, J., & Farhadi, A. (2017). *YOLO9000: better, faster, stronger*. Pages 7263-7271. Proceedings of the IEEE conference on computer vision and pattern recognition.
30. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You only look once: Unified, real-time object detection*. Pages 779-788. Proceedings of the IEEE conference on computer vision and pattern recognition.
31. Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). *Generalized intersection over union: A metric and a loss for bounding box regression*. Pages 658-666. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.
32. Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., & Wang, Z. (2016). *Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network*. Pages 1874-1883. Proceedings of the IEEE conference on computer vision and pattern recognition.
33. Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). *Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation*. Pages 1015-1021. Australasian joint conference on artificial intelligence: Springer.
34. Tagne, A., Feujio, T., & Sonna, C. (2008). *Essential oil and plant extracts as potential substitutes to synthetic fungicides in the control of fungi*. Pages 12-15. International Conference Diversifying crop protection.
35. Tompson, J., Goroshin, R., Jain, A., LeCun, Y., & Bregler, C. (2015). *Efficient object localization using convolutional networks*. Pages 648-656. Proceedings of the IEEE conference on computer vision and pattern recognition.
36. Ubbens, J., Cieslak, M., Prusinkiewicz, P., & Stavness, I. (2018). The use of plant models in deep learning: an application to leaf counting in rosette plants. *Plant methods*, *14*, 1-10. https://doi.org/10.1186/s13007-018-0273-z
37. Wang, C. Y., Liao, H. Y. M., Wu, Y. H., Chen, P. Y., Hsieh, J. W., & Yeh, I. H. (2020). *CSPNet: A new backbone that can enhance learning capability of CNN*. Pages 390-391. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops.
38. Xiong, H., Cao, Z., Lu, H., Madec, S., Liu, L., & Shen, C. (2019). TasselNetv2: in-field counting of wheat spikes with context-augmented local regression networks. *Plant Methods*, *15*, 1-14. https://doi.org/10.1186/s13007-019-0537-2
39. Zhu, M. (2004). *Recall, precision and average precision*. Department of Statistics and Actuarial Science, University of Waterloo, Waterloo 2: 6.
40. Zou, H., Lu, H., Li, Y., Liu, L., & Cao, Z. (2020). Maize tassels detection: a benchmark of the state of the art. *Plant Methods*, *16*, 1-15. https://doi.org/10.1186/s13007-020-00651-z